

# Combined spatial and tidal processes identify links between pelagic prey species and seabirds

S. L. Cox<sup>1,2,\*</sup>, B. E. Scott<sup>1</sup>, C. J. Camphuysen<sup>3</sup>

<sup>1</sup>School of Biological Sciences, Institute of Biological and Environmental Sciences, University of Aberdeen, Tillydrone Avenue, Aberdeen AB24 2TZ, UK

<sup>2</sup>School of Marine Science and Engineering (Faculty of Science and Technology), Plymouth University, Drake Circus, Plymouth PL4 8AA, UK

<sup>3</sup>Netherlands Institute for Sea Research (NIOZ), PO Box 59, 1790 AB Den Burg, Texel, The Netherlands

samantha.cox@plymouth.ac.uk

*Marine Ecology Progress Series: 479: 203–221 (2013)*

---

## Supplement 1. Principle component analysis (PCA) of nautical area scattering coefficient (NASC) threshold bands

Based on backscattering strengths, fisheries acoustic data were originally sorted into 4 threshold bands: greater than  $-40$  dB,  $-40$  to  $-50$  dB,  $-50$  to  $-60$  dB and  $-60$  to  $-70$  dB. Each threshold band contained NASC values over 5 min intervals for every 5 m change in depth through the water column. Corresponding to the time allocations of the observational 5 min seabird bins, acoustic data in the 4 threshold bands were allocated bin identifications. This resulted in each bin containing one rotation of acoustic density measurements at depth. For each threshold band, the maximum NASC value within each bin and the depth at which it occurred was extracted. This created 2 variables for each threshold band,  $NASC_{MAX}$  and  $Depth_{MaxNASC}$ , resulting in a total of 8 prey variables.

Multivariate datasets can be dealt with by rearranging the variance based on the association or correlation between variables, with the aim to reduce many variables to a smaller number of new derived variables that adequately describe the original data (Quinn & Keough 2006). This can be done using PCA.

To reduce the number of variables and account for any uncertainty surrounding threshold allocations, a PCA was run to distinguish similarities in the different threshold bands. At thresholds of less than  $-50$  dB, objects are likely to be similar because they represent weak scatterings. The PCA confirmed this (Table S1). Variables from the  $-50$  to  $-60$  dB and  $-60$  to  $-70$  dB threshold bands were found to explain much of the same variation in the data.

Rather than incorporate the new PCA variables into the dataset and lose information by combining  $NASC_{MAX}$  and  $Depth_{MaxNASC}$ , it was decided to create 2 new variables,  $NASC-50-70_{MAX}$  and  $Depth-50-70_{MaxNASC}$ . NASC values from the  $-50$  to  $-60$  dB and  $-60$  to  $-70$  dB threshold bands were summed to create a new threshold band,  $-50$  to  $-70$  dB. The resulting new  $NASC_{MAX}$  and corresponding  $Depth_{MaxNASC}$  were then extracted for each 5 min bin. Despite NASC being proportional to density, it is possible to sum values from different depths or threshold bands (MacLennan & Simmonds 1992). Values cannot be summed horizontally through the water column if they are from the same threshold band.

Table S1. Results from the principle component analysis (PCA) are shown. *Depth-40<sub>MaxNASC</sub>* was not included in the analysis due to a high number of NA measurements resulting from 0 values of *NASC-40<sub>MAX</sub>*. Values above 0.4 are marked with an asterisk (\*) indicating the variables represented by the component that are similar to one another

Variable	PCA 1
<i>NASC-40<sub>MAX</sub></i>	0.056
<i>Depth-40-50<sub>MaxNASC</sub></i>	-0.147
<i>NASC-40-50<sub>MAX</sub></i>	0.158
<i>Depth-50-60<sub>MaxNASC</sub></i>	-0.511*
<i>NASC-50-60<sub>MAX</sub></i>	0.462*
<i>Depth-60-70<sub>MaxNASC</sub></i>	-0.442*
<i>NASC-60-70<sub>MAX</sub></i>	0.529*
Importance of components	
Standard deviation	1.506
Proportion of variance	0.324
Cumulative proportion	0.324

## Supplement 2. Bird abundance mapped against oceanographic parameters

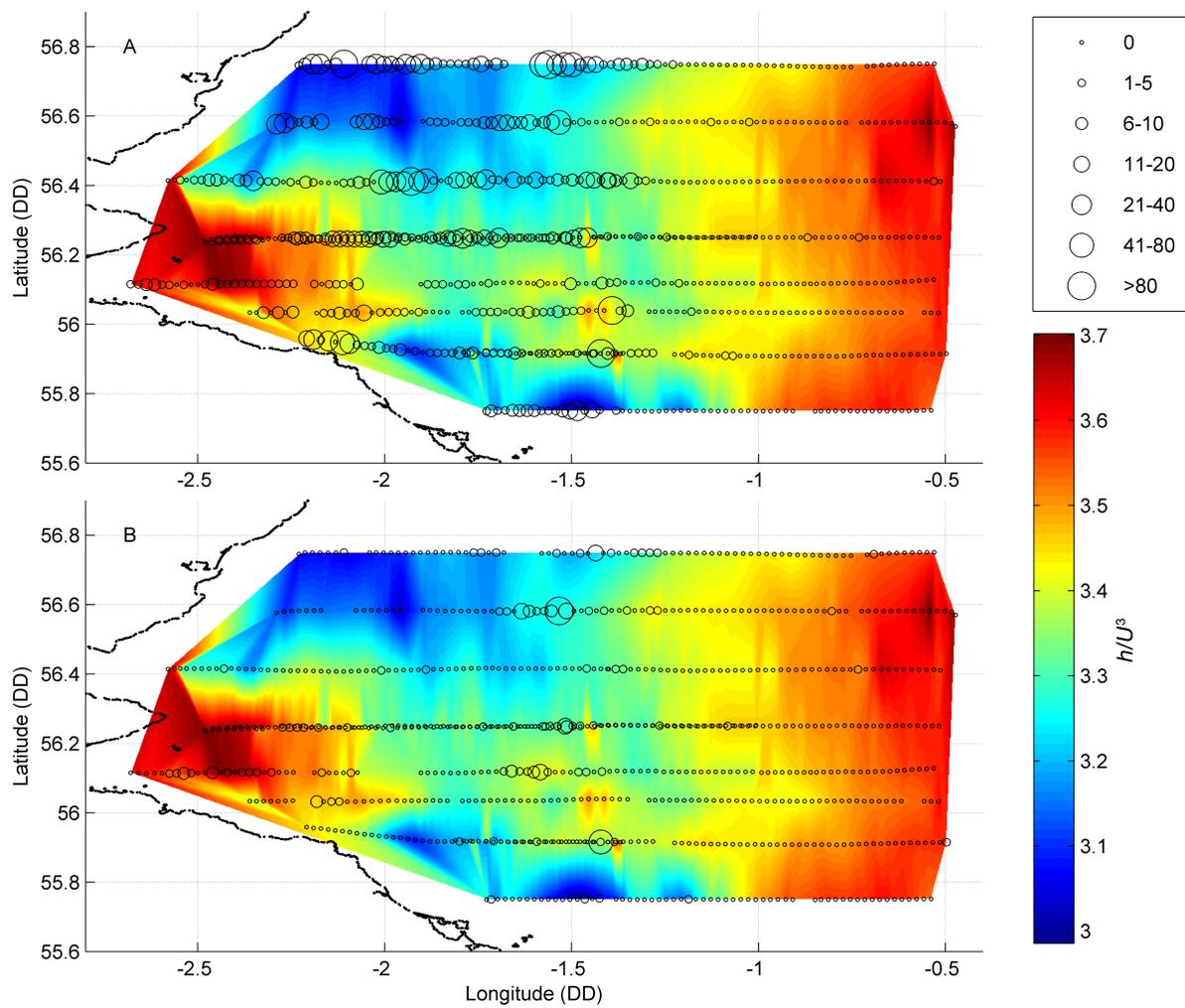


Fig. S1. *Uria aalge* and *Rissa tridactyla*. Distributions of the 2 seabird species in relation to  $h/U^3$ . (A) Guillemots, (B) kittiwakes. Values of  $h/U^3$  between the survey transects were calculated using a linear interpolation of summary statistics of the 5 min bins

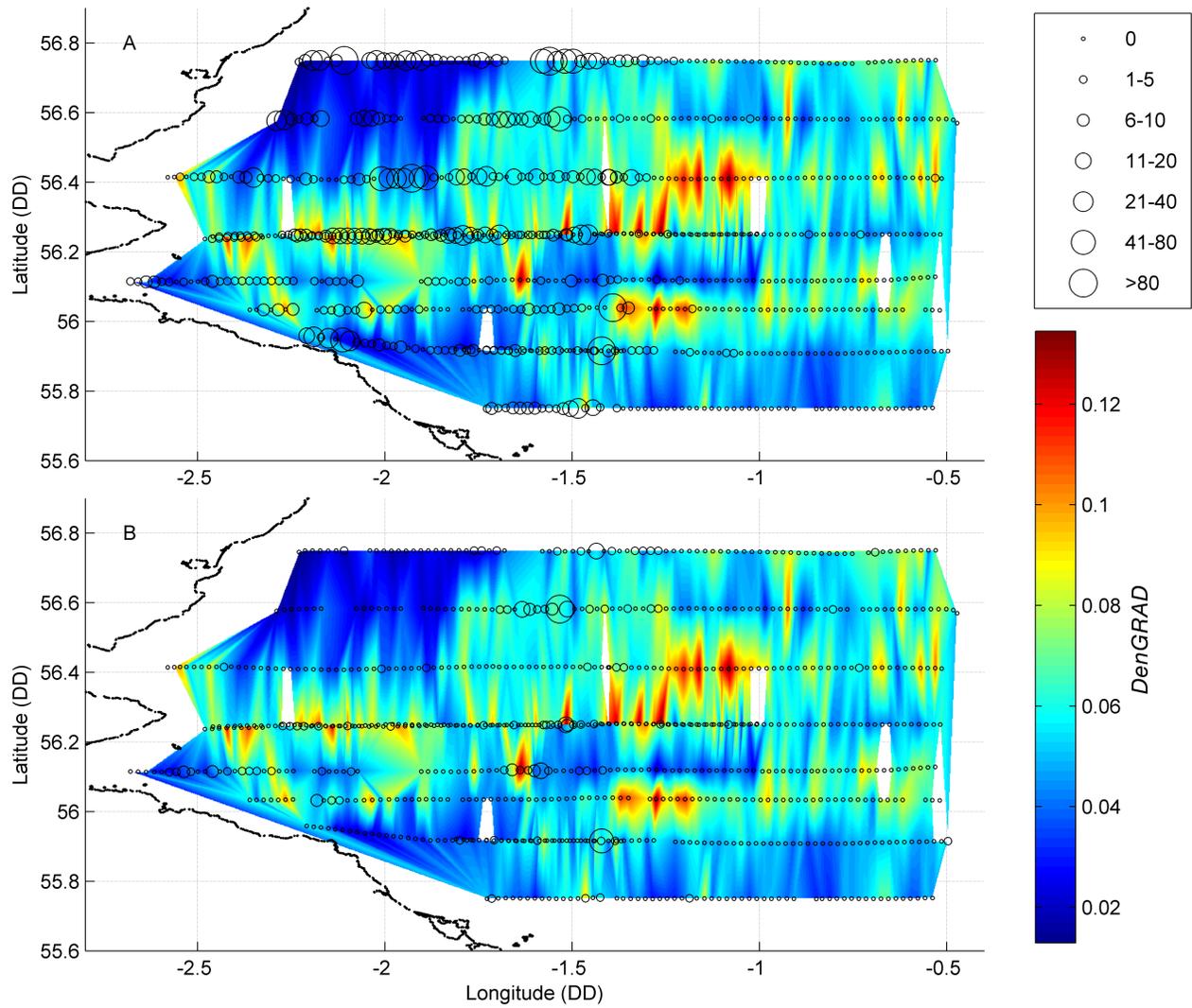


Fig. S2. *Uria aalge* and *Rissa tridactyla*. Distributions of the 2 seabird species in relation to *DenGRAD*. (A) Guillemots, (B) kittiwakes. *DenGRAD* between the survey transects was calculated using a linear interpolation of summary statistics of the 5 min bins

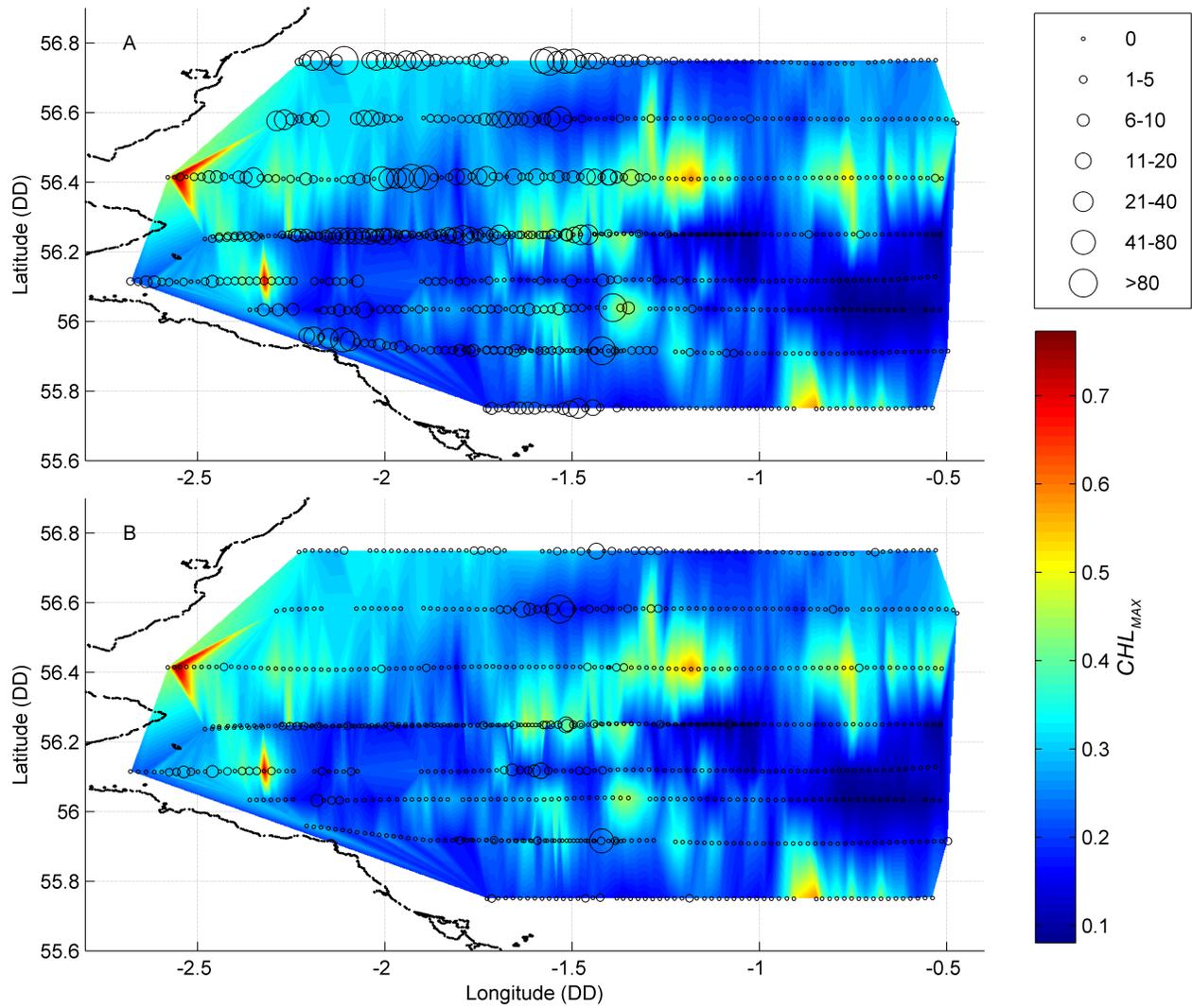


Fig. S3. *Uria aalge* and *Rissa tridactyla*. Distributions of the 2 seabird species in relation to  $CHL_{MAX}$ . (A) Guillemots, (B) kittiwakes.  $CHL_{MAX}$  between the survey transects was calculated using a linear interpolation of summary statistics of the 5 min bins

### Supplement 3. Prey densities mapped against oceanographic parameters

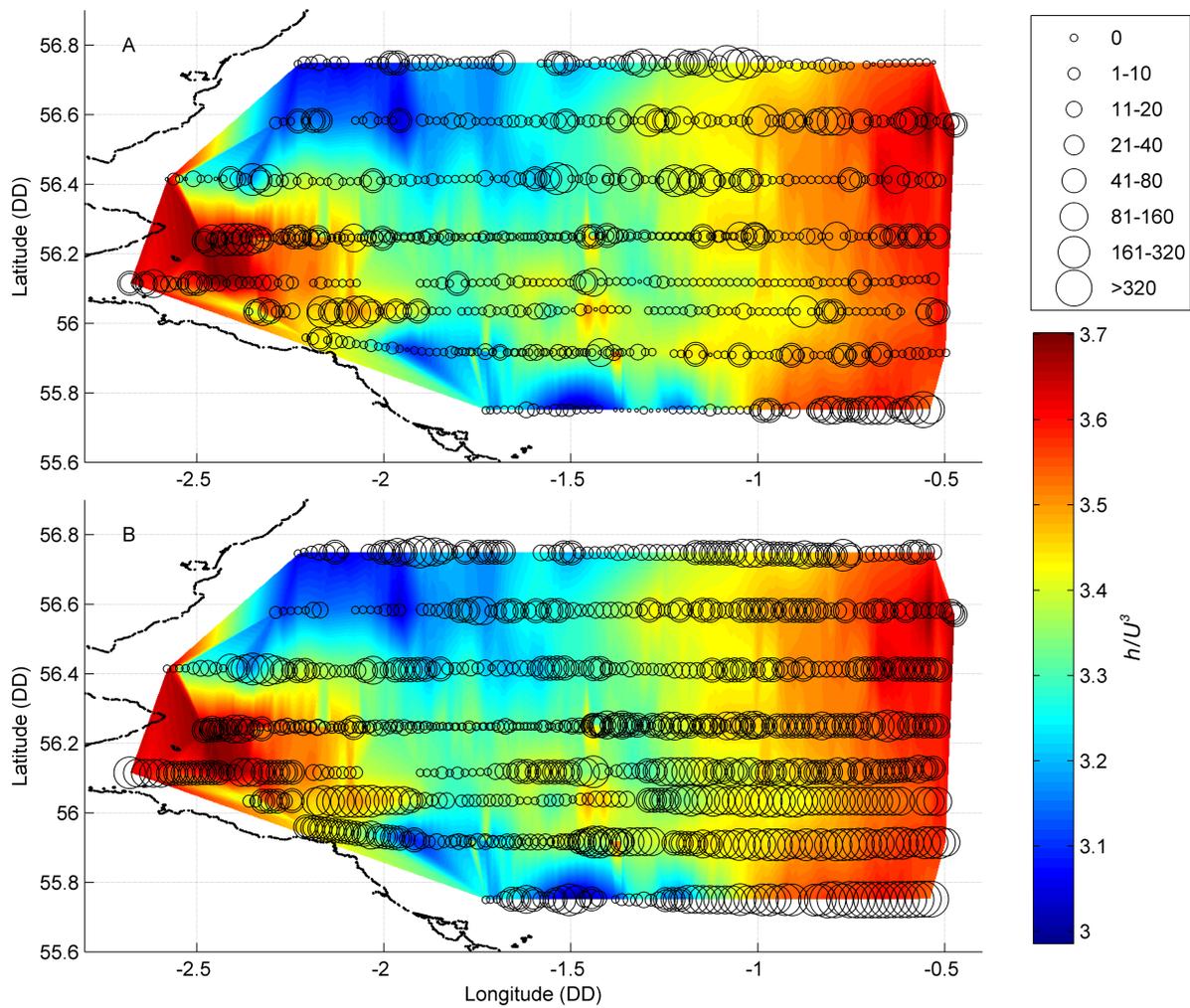


Fig. S4. Distributions of the 2 prey threshold bands in relation to  $h/U^3$ . (A)  $NASC-40-50_{MAX}$ , (B)  $NASC-50-70_{MAX}$ . Values of  $h/U^3$  between the survey transects were calculated using a linear interpolation of summary statistics of the 5 min bins

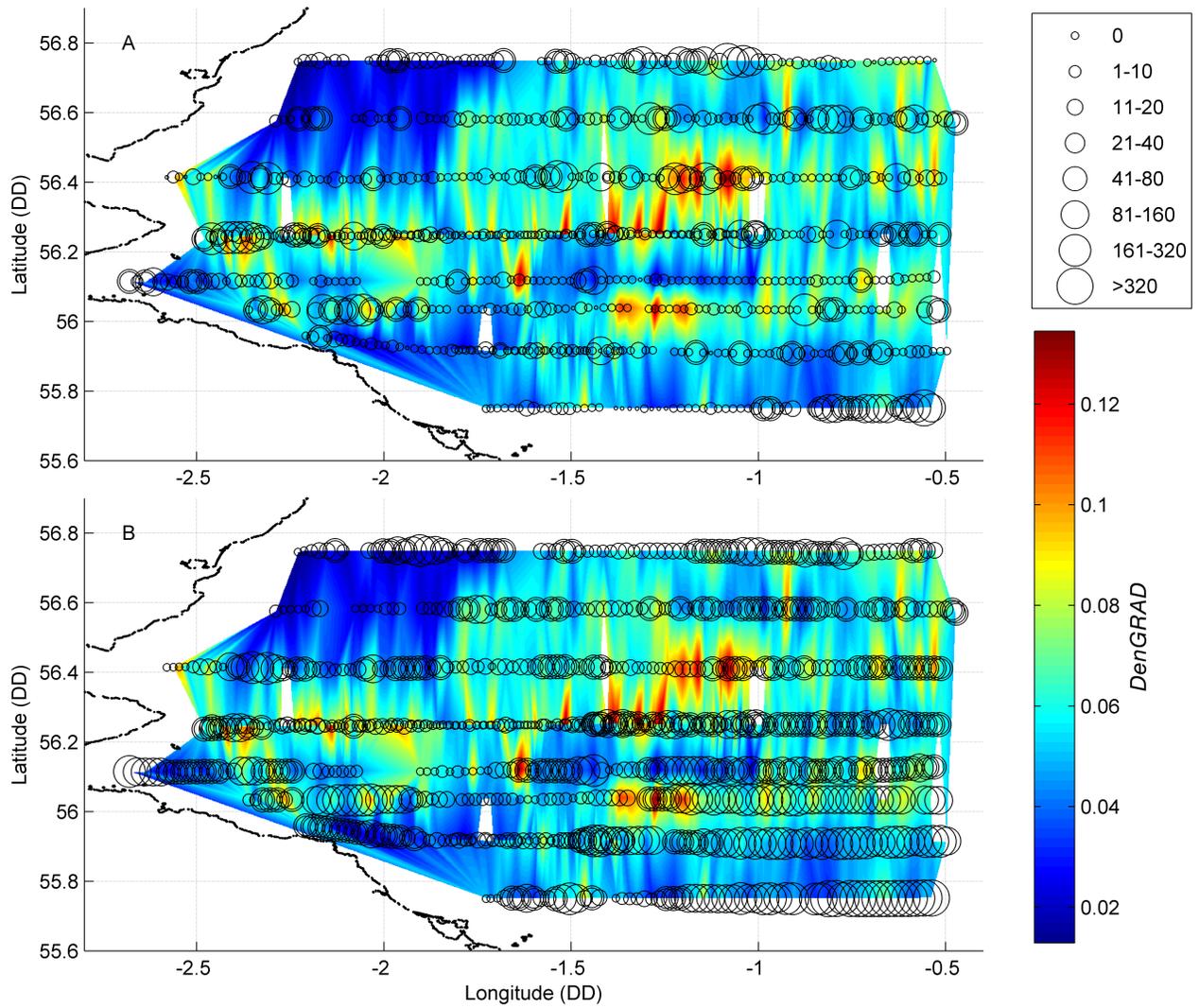


Fig. S5. Distributions of the 2 prey threshold bands in relation to *DenGRAD*. (A) *NASC-40-50\_MAX*, (B) *NASC-50-70\_MAX*. *DenGRAD* between the survey transects was calculated using a linear interpolation of summary statistics of the 5 min bins

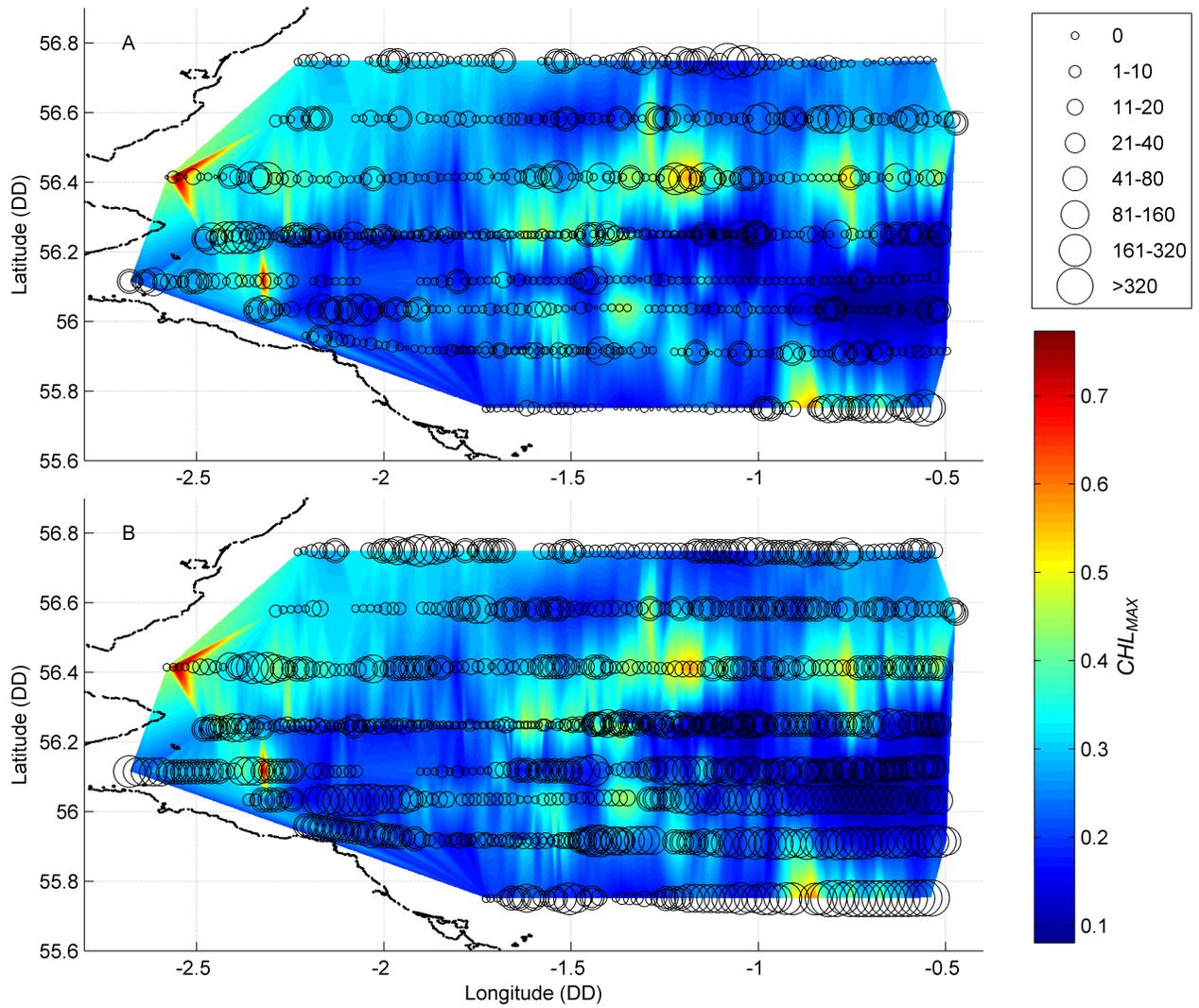


Fig. S6. Distributions of the 2 prey threshold bands in relation to  $CHL_{MAX}$ . (A)  $NASC-40-50_{MAX}$ , (B)  $NASC-50-70_{MAX}$ .  $CHL_{MAX}$  between the survey transects was calculated using a linear interpolation of summary statistics of the 5 min bins

#### Supplement 4. Full model selection results of seabird models

Table S2. *Uria aalge*. Full model selection results for both the habitat ( $\pi$ ) and count ( $\mu$ ) parts of the guillemot zero-inflated model. It is noted that  $SD(H)_{LOG}$  could be included in the model with borderline Akaike's information criterion (AIC) and p-values. However, the inclusion of this variable caused the model to break down where the model failed to converge on the removal of other variables. Furthermore, the variables included in the final model did not change with the inclusion of  $SD(H)_{LOG}$ . Therefore, given its borderline position, it was decided to exclude it from the model

Dropped term	df	AIC	Likelihood ratio test		
			Chi-squared	df	p
<i>None</i>	17	2864.593			
<i>Tidal Flag</i> from $\mu$	12	2958.793	104.200	5	<0.001
<i>WindSPEED</i> from $\mu$	16	2868.608	6.015	1	0.014
<i>WindDIR</i> from $\mu$	16	2868.953	6.361	1	0.012
<i>ThermSTRAT</i> from $\mu$	16	3013.869	151.280	1	<0.001
<i>Log NASC-40-50<sub>MAX</sub></i> from $\mu$	16	2868.714	6.121	1	0.013
<i>DenGRAD<sup>2</sup></i> from $\mu$	15	2889.737	29.144	2	<0.001
<i>Depth-50-70<sub>MaxNASC</sub></i> from $\mu$	16	2915.395	52.802	1	<0.001
<i>h/U<sup>3</sup></i> from $\pi$	16	2873.568	10.975	1	0.001
<i>CHL<sub>MAX</sub></i> from $\pi$	16	2909.173	46.580	1	<0.001
<b>Added term</b>					
<i>CHL<sub>MAX</sub></i> to $\mu$	18	2864.693	1.900	1	0.168
<i>Log NAS -40<sub>MAX</sub></i> to $\mu$	18	2864.596	7.997	1	0.158
<i>Log NASC-50-70<sub>MAX</sub></i> to $\mu$	18	2865.955	0.6373	1	0.425
<i>Depth-40-50<sub>MaxNASC</sub></i> to $\mu$	18	2867.34	0.747	1	0.387
<i>SD(H)<sub>LOG</sub></i> to $\pi$	18	2862.43	4.167	1	0.041

Table S3. *Rissa tridactyla*. Full model selection results for both the habitat ( $\pi$ ) and count ( $\mu$ ) parts of the kittiwake zero-inflated model

Dropped term	df	AIC	Likelihood ratio test		
			Chi-squared	df	p
None	13	919.080			
<i>Tidal Flag</i> from $\mu$	8	929.352	20.271	5	0.001
<i>Log NASC -40-50<sub>MAX</sub></i> from $\mu$	12	925.768	8.687	1	0.003
<i>ThermSTRAT<sup>2</sup></i> from $\mu$	11	932.522	17.441	2	<0.001
$(h/U^3)^2$ from $\pi$	11	921.188	6.108	2	0.047
<hr/>					
Added term					
<i>WindSPEED</i> to $\mu$	14	918.629	2.452	1	0.117
<i>WindDIR</i> to $\mu$	14	922.863	1.782	1	0.182
<i>DenGRAD<sup>2</sup></i> to $\mu$	16	922.290	0.790	2	0.674
<i>CHL<sub>MAX</sub></i> to $\mu$	14	919.684	1.396	1	0.237
<i>Log NASC-40<sub>MAX</sub></i> to $\mu$	14	919.198	1.882	1	0.170
<i>Log NASC-50-70<sub>MAX</sub></i> to $\mu$	14	920.532	0.548	1	0.459
<i>Depth-40-50<sub>MaxNASC</sub></i> to $\mu$	14	921.590	0.509	1	0.475
<i>Depth-50-70<sub>MaxNASC</sub></i> to $\mu$	14	921.492	0.411	1	0.521
<i>CHL<sub>MAX</sub></i> to $\pi$	14	920.931	0.150	1	0.699
<i>SD(H)<sub>LOG</sub></i> to $\pi$	14	922.825	1.745	1	0.187

## Supplement 5. Model validation

One extreme value was identified in the guillemot *Uria aalge* response variable. Therefore, the model was refit with the exclusion of this observation to determine its influence on results. No notable difference was observed, and similar relationships between explanatory and response variables were found.

Two extreme values were identified in the kittiwake *Rissa tridactyla* response variables. Refitting the model with the exclusion of these outliers did not make any notable difference to the results from the count part of the model. However, in the habitat part of the model, refitting the model with the exclusion of the 2 outliers meant that  $h/U^3$  was dropped and only an intercept was found to be significant in explaining the excess zero values. The increased sensitivity of the kittiwake model in comparison to the guillemot model may be due to the low number of positive observations, making it difficult for the model to determine how habitat covariates influence the probability of obtaining a zero observation. Therefore, the removal of just 2 positive observations may influence how this part of the model performs.

Despite being deemed acceptable, residual plots for both seabird models were not perfect and, therefore, some degree of caution should surround the interpretation of results. Residuals plotted against response variables were satisfactory. However, residuals plotted against explanatory variables and fitted values showed some evidence of heteroscedasticity, increasing the likelihood of rejecting a true null hypothesis (Type 1 error). However, given the consistency of results between species, with all 4 models (with and without the inclusion of extreme values) indicating a strong influence of thermal stratification and tidal speed/direction on driving both predator and prey distributions and both seabird species displaying similar responses to prey variables, conclusions can be made with some degree of confidence. Explanatory variables identified solely in the guillemot model (*WindSPEED*, *WindDIR* and *Depth-50-70<sub>MaxNASC</sub>*) warrant further research and, therefore, firm conclusions were not made. Despite this, however, an insight into additional driving factors of variability in guillemot abundance within the North Sea is provided.

Goodness of fit was evaluated by 2 methods. First, the fitted values of the model were plotted against the observed, and a linear model was then fit. Using this method, an intercept of 0 and slope of 1 would indicate a model of perfect calibration, where the observed values equalled the fitted values (Potts & Elith 2006). Both seabird models showed adequate goodness of fit. The second method was to plot a histogram of the residuals to observe normality. Both models showed satisfactory normality of the residuals.

Finally, a common concern with spatial data is correlation between observations, where the purpose of including a spatial correlation structure in a model is to account for similarity of counts between neighbours (Zuur et al. 2009). The influence of including a spatial correlation structure could not be determined, as general software code for R is not currently available and still in developmental stages (Zuur et al. 2009). However, often the reason for such correlations is a result of similarities in underlying habitat features within a neighbourhood (Zipkin et al. 2010). If birds are indeed responding to such habitat features, including them within the model should account for patterns in abundance variation, making a spatial correlation structure redundant (Dormann 2007). Indeed, in contrast to residual plots from the prey models (where a spatial correlation structure was needed), residuals of both the seabird models plotted through space did not show obvious signs of grouping, which would indicate an underlying issue of potential non-independence. Furthermore, Embling et al. (2012) showed that increases in kittiwake abundance from a sister dataset occurred over distances of 2.5 km, suggesting that any correlations would not exceed this distance (2 of our 5 min bins). Therefore, the subsequent influence would be small because the vast majority of comparisons would not be correlated. Indeed semi-variograms suggested that any correlation issues present were highly localised and, therefore, were likely to have little impact on results.

**Supplement 6. Prey model selection**

Table S4. Model selection of explanatory variables for *Log NASC-40-50<sub>MAX</sub>*

Dropped term	df	AIC	BIC	Likelihood ratio test		
				L	df	p
<i>None</i>	10	2365.076	2411.847			
<i>Tidal Flag</i>	5	2368.838	2392.223	13.763	5	0.017
<i>ThermSTRAT</i>	9	2380.137	2422.230	17.061	1	<0.001
<b>Added term</b>						
<i>DenGRAD</i>	11	2365.620	2417.068	1.456	1	0.228
<i>h/U<sup>3</sup></i>	11	2365.894	2417.341	1.182	1	0.277
<i>CHL<sub>MAX</sub></i>	11	2367.071	2367.071	0.005	1	0.944
<i>SD(H)<sub>LOG</sub></i>	11	2364.194	2364.194	2.882	1	0.090

Table S5. Model selection of explanatory variables for *Log NASC-50-70<sub>MAX</sub>*

Dropped term	df	AIC	BIC	Likelihood ratio test		
				L	df	p
<i>None</i>	10	1086.729	1133.500			
<i>Tidal Flag</i>	5	1148.270	1171.655	71.540	5	<0.001
<i>ThermSTRAT</i>	9	1093.612	1135.706	8.883	1	0.003
<b>Added term</b>						
<i>DenGRAD</i>	11	1088.729	1140.177	0.000	1	0.992
<i>h/U<sup>3</sup></i>	11	1088.177	1139.624	0.553	1	0.457
<i>CHL<sub>MAX</sub></i>	11	1088.691	1140.139	0.038	1	0.845
<i>SD(H)<sub>LOG</sub></i>	11	1086.430	1137.878	2.299	1	0.129

## LITERATURE CITED

- Dormann CF (2007) Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Glob Ecol Biogeogr* 16:129–138
- Embling CB, Illian J, Armstrong E, Van der Kooij J, Sharples J, Camphuysen CJ, Scott BE (2012) Investigating fine-scale spatio-temporal predator–prey patterns in dynamic marine ecosystems: a functional data analysis approach. *J Appl Ecol* 49:481–492
- MacLennan DN, Simmonds EJ (1992) *Fisheries acoustics*. Chapman & Hall, London
- Potts JM, Elith J (2006) Comparing species abundance models. *Ecol Model* 199:153–163
- Quinn GP, Keough MJ (2002) *Experimental design and data analysis for biologists*. Cambridge University Press, Cambridge
- Zipkin EF, Gardner B, Gilbert AT, O’Connell AF Jr, Royle JA, Silverman ED (2010) Distribution patterns of wintering sea ducks in relation to the North Atlantic Oscillation and local environmental characteristics. *Oecologia* 163:893–902
- Zuur AF, Ieno EN, Walker NJ, Saveliev AA, Smith GM (2009) *Mixed effects models and extensions in ecology with R*. Springer Verlag, New York, NY